Advanced Review

The power of coarse graining in biomolecular simulations

Helgi I. Ingólfsson, Cesar A. Lopez, Jaakko J. Uusitalo, Djurre H. de Jong, Srinivasa M. Gopal, Xavier Periole and Siewert J. Marrink*

Computational modeling of biological systems is challenging because of the multitude of spatial and temporal scales involved. Replacing atomistic detail with lower resolution, coarse grained (CG), beads has opened the way to simulate large-scale biomolecular processes on time scales inaccessible to all-atom models. We provide an overview of some of the more popular CG models used in biomolecular applications to date, focusing on models that retain chemical specificity. A few state-of-the-art examples of protein folding, membrane protein gating and self-assembly, DNA hybridization, and modeling of carbohydrate fibers are used to illustrate the power and diversity of current CG modeling. © 2013 John Wiley & Sons, Ltd.

How to cite this article:
WIREs Comput Mol Sci 2014, 4:225–248. doi: 10.1002/wcms.1169

INTRODUCTION

To unveil the driving forces governing biomolecular processes, computer simulations have become an remarkable tool, in particular, the molecular dynamics (MD) simulation technique. In this approach, the time evolution of a system of interacting particles is computed mainly based on pairwise forces between the atoms. MD simulations have advanced to such a level that nowadays one can talk about 'computational microscopy' as an added tool to experimental microscopy methods.[1,2] Traditional all-atom models are inadequate to simulate the large spatiotemporal scales involved in cellular processes. Coarse grained (CG) models have gained a lot of popularity lately as by neglecting some of the atomistic degrees of freedom (DOFs) they allow for a significant increase over both the spatial and temporal limitations of all-atom models.[3,4]

One of the current challenges in the field of biomolecular modeling is to develop accurate and transferable CG force fields (FFs). Essentially, two different routes are followed. In bottom-up approaches (structure-based coarse graining), effective CG interactions are extracted from reference atomistic simulations. This can be done in a systematic way by using inverse Monte Carlo (IMC),[5] iterative Boltzmann inversion (IBI),[6] force matching (FM),[7,8] or related methods (see Box 1). In top-down approaches (thermodynamic-based coarse graining), the focus is on reproducing key experimental data, especially thermodynamic properties. Typically, simple analytical interaction potentials are used and the parameters are optimized in an iterative procedure. Although the bottom-up approaches are capable of capturing more of the fine details of the interaction, the top-down approach usually provides potentials that are more easily transferable. In practice, many CG FFs rely on a combination of these two routes. For a recent review on different approaches to coarse graining, see Brini et al.[9]

Provided that one respects the limitations of the CG model at hand, CG modeling has five powerful advantages, namely: (1) enabling efficient simulations of huge system sizes, with simulation volumes up to

The authors have declared no conflicts of interest in relation to this article.

*Correspondence to: s.j.marrink@rug.nl

Groningen Biomolecular Sciences and Biotechnology Institute & Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands

DOI: 10.1002/wcms.1169

BOX 1: STRUCTURE-BASED COARSE GRAINING

In structure-based coarse graining, CG potentials are constructed in such a way that predefined target functions, which structurally characterize the system, are reproduced in the CG simulation. The target functions are mostly obtained from higher resolution atomistic simulations, but in more knowledge-based approaches, experimentally derived structural data can also be used.

In the commonly used IBI method,[6] radial distribution functions, $g_{\text{ref}}(r)$, are the target reference functions. Through the simple Boltzmann inversion

$$V_{\text{PMF}}(r) = k_{\text{B}}T \ln[g_{\text{ref}}(r)], \tag{1}$$

where k_{B} denotes the Boltzmann constant and T the temperature, the potential of mean force (PMF) V_{PMF} between pairs of CG particles as a function of their distance r can be obtained. Unfortunately, this PMF cannot be directly used as a pair potential in a CG model because it encloses multibody contributions from all the particles in the system. Therefore, an iterative procedure should be used to extract the intermolecular CG potential V^{CG}:

$$V^{\text{CG}}_{i+1}(r) = V^{\text{CG}}_{i}(r) + k_{\text{B}}T \ln\left[\frac{g_i(r)}{g_{\text{ref}}(r)}\right] \tag{2}$$

The procedure is initiated with V_{PMF} extracted from the simple Boltzmann inversion. The subscript i denotes the iteration number. According to the Henderson theorem,[13] the IBI method guarantees the theoretical uniqueness of the two-body CG interaction potential for the given $g_{\text{ref}}(r)$.

FM[7,8] is another popular technique for constructing CG potentials. FM does not rely on pair correlation functions, that is, pair potentials of mean force, but instead matches forces on the CG interaction sites as closely as possible with the forces at the atomistic level. Thus, it aims at reproducing the multibody PMF with a set of CG interaction functions. In order to determine FM CG potentials, first reference forces $\boldsymbol{F}_i^{\text{ref}}$ on CG beads are calculated as a sum of the associated atomistic forces f_γ

$$\boldsymbol{F}_i^{\text{ref}} = \sum_{\gamma} f_\gamma \tag{3}$$

Next, a model is constructed in which the CG FF depends linearly on a number of fitting parameters, the coefficients of cubic splines used to tabulate the CG forces. Subsequently, a fitting procedure is performed, which in essence involves a solution to the following set of $N \times L$ equations,

$$F^{\text{CG}}_{il}(g_1 \ldots g_m) = F^{\text{ref}}_{il}, \quad i = 1, \ldots, N; l = 1, \ldots, L \tag{4}$$

where $g_1 \ldots g_m$ are the fitting parameters, N is the number of CG beads, and L is the number of reference frames used for the coarse graining. L should be large enough to make the set of equations overdetermined. The calculation is usually repeated for a number of smaller parts of the trajectory and the final result is constructed as an average over the set of solutions.

A number of other structure-based methods exist, for example, IMC,[5] minimization of relative entropy,[14] or conditional reversible work,[15] but so far IBI and FM have been most widely used in the development of effective CG potentials for biomolecular simulations. For an excellent overview, see the work of Noid.[16]

It is important to realize that there is no unique method to construct CG potentials from higher resolution data. A full representation of higher order correlations requires multibody potentials, which are impractical and computationally expensive thereby defeating the purpose of coarse graining. Even when the pair correlations are well described, other system properties such as the pressure or energy cannot be matched at the same time. The art of coarse graining is in the compromise of assessing which level of detail needs to be included. In the end, the most suitable CG method depends on the type of questions asked.

$100 \times 100 \times 100\,\text{nm}^3$ containing millions of particles; (2) allowing for the simulation of slow processes requiring time scales in the micro- to millisecond range; (3) enabling high-throughput studies, systematically exploring state conditions in thousands of parallel runs; (4) showing where details matter and where not when compared with higher resolution methods, thus providing insights into the physical nature of the fundamental driving forces; and (5) providing a computationally inexpensive testing ground for exploring novel generic biophysical pathways. Compared with all-atom models, CG models are easily two to three orders of magnitude faster, as a result of fewer DOFs combined with larger integration time steps and faster sampling due to a smoothened energy landscape (see Box 2).

Here, we review some of the more successful methods that have been developed for large-scale simulation of biomolecular systems, as well as highlight some novel and promising approaches. To narrow the scope, we restrict ourselves to particle-based simulation methods and only describe models that retain chemical specificity. These models present the advantage of providing an easy access to full atomistic details using resolution transformation methods.[10–12] This review is organized as follows. First we give a survey of CG models grouped according to class of biomolecule. We then provide a number of

BOX 2: WHY ARE CG MODELS SO FAST?

The major incentive to use CG models is the fast sampling they provide. There are four reasons for this speedup:

1. *Reduced number of DOFs.* For CG models that retain chemical specificity, the typical reduction factor in the number of particles, n, is between 3 and 5 with respect to united atom FFs, or around 10 compared with fully atomistic FFs that include explicit hydrogens. For models that map multiple water molecules into a single CG bead, the solvent DOFs are also greatly reduced. Fewer particles to compute and fewer neighbors to consider cause a speedup of the order of n^2. Taking the Martini model as an example, with $n = 4$ for most molecules compared to a united atom model and $n = 12$ for water, a speedup of 16–144 is obtained depending on overall water content. In models where the solvent is omitted entirely, the speedup is orders of magnitude larger.
2. *Short range interactions.* Most CG models only compute short-range interactions, typically cutoff at a distance around 1 nm. No expensive PME methods are needed as the electrostatic interactions are effectively captured in the short-range potentials. Furthermore, many CG models shift the potentials to zero at the cutoff, allowing for less-frequent pairlist updates. Compared with a typical setup for atomistic simulations with full treatment of long-range electrostatics, an order of magnitude speedup is easily obtained.
3. *Faster dynamics.* Because of the loss of atomistic DOFs, the potential energy surface is smoothed out leading to reduced friction. In the same simulation time, a CG system can therefore sample more of the phase space. The speedup factor is difficult to generalize because the amount of friction removed depends quite sensitively on the nature of the mapping. H-bonds, for instance, are likely to contribute much more to the friction than methylene groups do.
4. *Larger integration time steps.* The overall smoother energy surface permits the use of larger integration time steps. Typical time steps used are tens of femtoseconds for MD, and >100 femtoseconds for dissipative particle dynamics (DPD), compared with 1–4 femtoseconds employed in all-atom MD simulations. However, CG models that used more detailed potentials obtained through IBI or FM, for instance, are limited to shorter time steps.

All together, the combined speedup factors are between 2 and 5 orders of magnitude for most of the CG models considered in this review. There is however no universal rule to predict the particular speedup of a CG model from the approximations and strategies it is based on. Consequently, the interpretation of time is problematic in CG models. The time scale is best calibrated by directly comparing with experimental data or dynamics from atomistic simulations for the system at hand.

state-of-the-art examples to illustrate the power of CG simulations. We conclude with a short outlook.

CG MODELS

In this section, we discuss the most widely used CG models as well as some promising recent methods. For each model, a brief characterization is given of the main features and limitations. The models are grouped according to the class of biomolecule, starting with water, followed by lipids, proteins, nucleotides, and carbohydrates. We do not claim, nor aim, to be exhaustive and we apologize if we have overlooked important contributions.

CG Water Models

Water is present as solvent in all biological systems and as such it is an important molecule to consider when developing biomolecular FFs. In spite of water's simple molecular structure, it shows in many aspects very complex collective behavior, making it a challenging molecule to model on a CG level. As a result, CG models have employed widely differing approaches to map (Figure 1) and parameterize water. The water models are divided into three categories: implicit, explicit, and polarizable models. As the current work deals with biomolecules, we focus on water models that have been used with or are suitable to be used in combination with a CG biomolecular FF.

Implicit Models

Roughly two main strategies have been used to model the aqueous phase implicitly. In a conceptually simple strategy, the hydrophobic effect and charge screening are accounted for by adjusting the nonbonded interactions between nonsolvent molecules. This strategy was applied by top-down (lipid) models,[23] protein folding FFs such as UNRESS,[24] the DNA model by Ouldridge et al.,[25] and the carbohydrate model by Srinivas et al.[26] In a more sophisticated strategy, the hydrophobic effect is modeled by a FF term depending

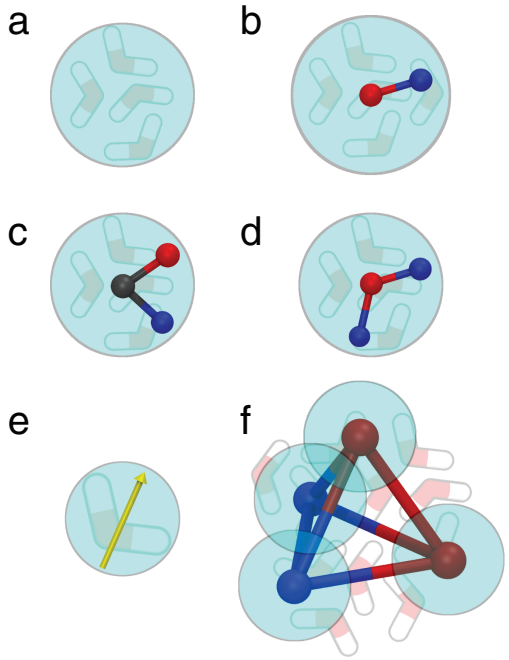

FIGURE 1 | Mapping strategies for CG water models: (a) regular Martini water,[17] (b) GROMOS CG water,[18] (c) Martini compatible polarizable water (PW),[19] (d) big multipole water (BMW),[20] (e) ELBA[21] induced point dipole water, and (f) Wat Four (WT4).[22]

on solute size and charges are explicitly screened using either Debye–Hückel theory or Generalized-Born models. Because of the emphasis on electrostatics, these methods have been popular in DNA models.[27,28] Note that although most of the implicit solvent models use an implicit representation for ions, some models combine implicit solvent with explicit ions.[29] Analogous to the screening of charges, some FFs scale the van der Waals (vdW) interactions using a term based on solvent accessible surface area.[30]

Explicit Models

An excellent overview of explicit CG water models is given in recent reviews by Hadley and McCabe[31] and Darré et al.[32] Explicit water models can be divided into models that have been parameterized using either structure or thermodynamics-based methods.

Structure-based models are constructed from atomistic simulations of water using methods such as IBI or FM (see Box 1). Several systematic studies looking at the properties and transferability of these water models are available.[33–35] As these methods rely on atomistic simulation trajectories, they have traditionally mapped one water molecule to one CG bead to avoid the challenging issue of grouping water molecules. This 1:1 mapping has limited the speedup of these models but allowed quite accurate CG water models to be conceived. The grouping issue has recently been circumvented by assembling together water molecules in the atomistic trajectory using a K-means algorithm[35] or the CUMULUS method.[36]

Thermodynamics-based models are parameterized by fitting to experimental solvent properties such as density, water–air surface tension, diffusion rates, or solvation free energies. These models use analytical potentials, most commonly Morse, Lennard-Jones (LJ), or Mie ones, and apply different mappings (Figure 1). He et al.[37] performed a systematic study of the properties of different potentials. A popular example of a thermodynamic-based water model is the water model associated with the Martini FF[17] (Figure 1a). The model represents four water molecules by a single CG bead (4:1 mapping) using a shifted LJ potential for the nonbonded interactions. It has been parameterized based on the density of pure water and the solubility of water in apolar solvents. An alternative water model compatible with the Martini FF uses a Morse potential for the nonbonded interactions.[38] The M3B model,[39] developed in connection with carbohydrates, is based on a 1:1 mapping and also uses a Morse potential. It was parameterized against the experimental density, intermolecular energy, and the diffusion coefficient of water. The thermodynamic-based water models used by Shelley et al.[40] and later by Shinoda et al.[41] use a 3:1 mapping and a 6–4 or 12–4 Mie potential, respectively. The first one was parameterized against water density, whereas the latter used density, compressibility, and air–water surface tension. The monoatomic water (mW) model by Molinero et al.[42] uses a nonbonded potential with two and three body terms mapping one water to one bead. This model reproduces the tetrahedral organization of water molecules in addition to a range of other properties such as density and phase transition temperatures.

The lack of charges in these explicit water models prevents them from screening electrostatics. Charge–charge interactions are either ignored or implicit screening is used.[42,43] For instance, the Martini model uses an implicit dielectric constant $\varepsilon = 15$. Ions, however, are typically included explicitly, and several of the CG water models have specific parameters available for ions.[22,43,44] The models applying a coarser mapping (e.g., Martini and WT4—see below) represent an ion together with its first solvation shell in one bead.

Polarizable Models

To alleviate the lack of proper electrostatic screening in the explicit water models discussed above, a number of polarizable CG water models have recently

been developed. Different methods have been used to mimic the electrostatic screening of water molecules, which arises in large part from their orientational polarizability. The most common approach is the introduction of extra particles carrying a charge. Two recent water models specifically aimed to be compatible with the Martini FF have been developed. Yesylevskyy et al.[19] proposed a model with two additional particles carrying a charge and bound to the LJ interaction site (Figure 1c). The relative rotation of the particles within a molecule—their polarization—is controlled by an angle potential and their interactions with the environment. Wu et al.[20] introduced the big multipole water (BMW) model (Figure 1d), in which water consists of three sites connected in a rigid V-shape. All three sites carry a partial charge, whereas only the central site is involved in vdW interactions—via a modified Born–Mayer–Huggins potential—with other water beads.

Several CG FFs model the polarizability of water using an induced point dipole. Examples are the ELBA FF parameterized by Orsi and Essex[21] (Figure 1e) and the polarizable pseudo-particle (PPP) model obtained by Ha-Duong et al.,[47] based on a (roughly) 1:1 mapping. For the PPP model, the induced dipoles are only susceptible to charges on other nonwater molecules. Quite different in topology is the Wat Four (WT4) model[22] (Figure 1f) that consists of four vdW spheres in a tetrahedral geometry that together map 11 water molecules. The beads interact via harmonic bonds and all four carry a charge. In addition to the original use in combination with a CG DNA model,[22] WT4 has also been used in multiscale simulations.[48] Other examples of CG water directed at multiscale simulations are the GROMOS CG water model[18] (Figure 1b), to be combined with the atomistic GROMOS FFs, and the PPP model in combination with the atomistic polarizable TCPEp FF.[49]

CG Lipid Models

Lipid bilayer structure and function has been extensively studied using all scales of molecular simulations. Although a typical lipid is rather small, around a hundred atoms, the bulk material properties of a lipid bilayer depend on the collective behavior of hundreds, if not hundreds of thousands, of lipids—rendering atomistic lipid bilayer simulations computationally costly. The large time and length scales required to study many of the interesting membrane-associated processes, such as lipid domain formation, sorting and clustering of membrane proteins, vesicle fusion and fission, and so on, have spurred the development of a large number of CG lipid FFs. The first CG lipid model dates back to 1990, by Smit et al.[50] Today, CG lipid models range all the way from continuum or semi-continuum models to atomistic or united atom models. Here, we focus on models that retain chemical specificity and are therefore able to distinguish specific lipid types. These kinds of models usually group 3–6 heavy atoms per CG bead, reducing a typical lipid to around 8–14 beads. This CG lipid mapping (Figure 2) is quite common and gives a good reduction in the number of particles and still allows enough flexibility for chemical specificity. Because of the large number of relevant CG methods available here, we only discuss recent models; for additional information please see recent reviews on CG lipid simulations.[23,51–54]

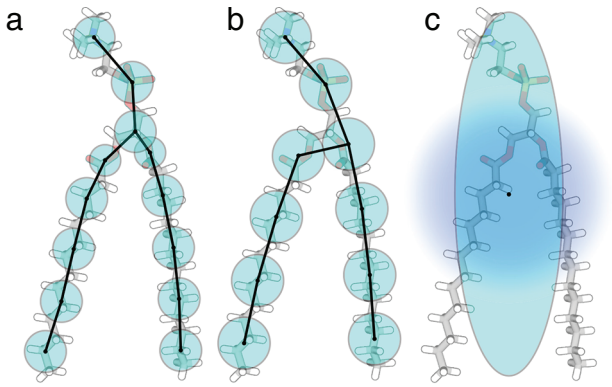

FIGURE 2 | Mapping strategies for CG lipid models illustrated for a dimyristoylphosphatidylcholine (DMPC) lipid. (a) Shelley et al.[40] and (b) Martini[17,43,45] models are overlaid on the atomistic structure. (c) The one bead per lipid aggressively CG model of Ayton and Voth[46] showing the analytical Gay-Berne ellipsoid particle model combined with an in-plane potential systematically derived from atomistic simulations.

Klein Models

Klein and coworkers are one of the pioneers in exploring CG lipid models. In Shelley et al.,[40] they demonstrated the feasibility of constructing a specific CG lipid FF directly from atomistic simulation data using dimyristoylphosphatidylcholine (DMPC) as an example. Each DMPC lipid is represented by 13 CG beads (Figure 2a) linked together using harmonic bond and quartic angle potentials, each one fitted to the underlying atomistic simulation. The nonbonded interactions were based on the radial distribution functions of the corresponding atomistic groups and refined using IBI. The resulting CG model reproduced the structural details of the atomistic simulations quite accurately but has limited transferability. There have been quite a few refinements to this model and other CG models have also been introduced,

including a promising new lipid model by Shinoda et al.[55] This new model uses the same atomistic to CG mapping scheme but CG particles representing the same atomistic group in different molecules are fitted jointly based on thermodynamic properties and multiple atomistic simulations. This strategy—combining bottom-up and top-down parameterization—resulted in improved transferability of the FF, both for different molecular structures and environmental conditions. Recent applications of this new model include studies of the phase behavior of lipid monolayers[55] and membrane partitioning of fullerenes.[56]

The Martini Model

The Martini FF was originally developed for lipids.[17,43,45] The philosophy behind Martini was not to capture every detail of a given atomistic simulation, but rather to present an extendable CG model based on simple modular building blocks, using few parameters and standard interaction potentials to maximize applicability and transferability. Martini uses an approximate 4:1 mapping (Figure 2b) and in version 2.0,[43] 18 bead types were defined to represent different levels of polarity as well as charged groups. The CG beads have a fixed size and interact using an interaction map with 10 different strengths. Both vdW and electrostatic interactions are described using shifted potentials and the electrostatics is screened with a relative dielectric constant $\varepsilon = 15$ using the standard Martini water or $\varepsilon = 2.5$ using the polarizable water model, see water section. Bonds and bond angles are described with harmonic potentials. Parameters were tuned to match thermodynamic and structural data from experimental as well as atomistic simulations of a number of systems. Because of the modularity of Martini, a large set of different lipid types has been parameterized (e.g., Refs 43, 45, 57, and 58) with applications ranging from vesicle self-assembly[45] to formation of raft domains[59] and membrane tethers[60] to name only a few.

The ELBA Model

The ELBA (electrostatics-based) CG lipid FF developed by Orsi and coworkers,[21,61] focuses on modeling lipid–water interactions and capturing important electrostatic contributions. The model represents each water molecule individually using soft sticky dipole potentials and incorporates electrostatics in the CG lipid beads as point charges or point dipoles—allowing for a relative dielectric constant $\varepsilon = 1$. A few lipid types have been parameterized[21] by matching lipid properties, such as volume and area per lipid, average segmental tail order parameter, spontaneous curvature, and dipole potential. Additionally, the ELBA FF was constructed with possible multiscale applications in mind.[62] Applications of the ELBA FF have thus far been focused on lipid phase behavior[21] and permeation of drugs and other compounds across bilayers.[62]

Voth Models

Voth and coworkers have developed numerous CG lipid models, for example, Refs 63–65 that, like the Klein model,[40] build the FFs directly from atomistic simulations, but instead of matching average structural properties, they target the underlying forces at the atomistic scale (FM, see Box 1). A typical atomistic lipid is mapped onto 13–15 CG beads, similar to the other models discussed in this section. Depending on the model, electrostatic interactions are treated explicitly or implicitly—by combining them with the short-range nonbonded potentials. Different models also represent water differently: explicitly incorporate each water molecule in one CG bead[63] or implicitly include the water contribution in the nonbonded potentials.[65] These methods have been demonstrated on a number of lipids, for example, DMPC,[63] cholesterol,[64] dioleoylphosphatidylcholine (DOPC), and dioleoylphosphatidylethanolamine.[65] By essence, this approach builds potentials that are difficult to transfer from one system to another.

Smit's DPD Model

Kranenburg et al.[66] studied CG lipid models using soft interaction potentials in DPD simulations. They modeled DMPC on two different CG scales: a fine (close to united atom scale) with a CG bead volume of 30 Å^3 and a 1:1 mapping for water and a coarser scale (13 beads for a DMPC lipid) with a CG bead volume of 90 Å^3 and a 3:1 mapping for water. In both models, the CG beads are connected with harmonic bonds and the bending potentials adjusted to fit distributions obtained from atomistic simulations. The DPD repulsion parameter set was determined by testing parameter combinations from various related DPD studies. The coarser model is quite fast even compared with other CG lipid models because of the very soft nature of the DPD potentials. The model has been improved upon a number of times and has been shown to describe the phase behavior for a variety of phospholipids and cholesterol quite accurately.[52,67–69]

Other Promising Models

A number of other models should be mentioned, in particular, recent attempts to parameterize solvent-free lipid models that retain chemical detail. Implicit solvent models gain considerably on

computation cost but do need to incorporate the excluded hydrophilic–hydrophobic interactions into the effective potentials between the CG beads. The models of Lyubartsev,[70] Wang and Deserno,[71] Sodt and Head-Gordon,[72] and Curtis and Hall[73] use similar number of CG beads per lipid (10–15) and derive their CG potentials from representative atomistic simulations. Wang and Deserno[71] and Sodt and Head-Gordon[72] add long-range attractive interactions on the lipid tails to mimic the hydrophobic effect, which they tune to fit experimental data. Curtis and Hall[73] in their LIME (Lipid Intermediate resolution ModEl) FF use hard-sphere and square-well potentials in order to use discontinuous molecular dynamics and gain even greater speedup.

Additionally, the Voth group introduced two supraresolution solvent-free methods: the Ayton and Voth[46] hybrid analytic-systematic (HAS) approach and the Srivastava and Voth[74] hybrid CG (HCG) models. These methods are aimed at even larger time and length scales, with applications such as modeling of large liposomes consisting of tens of thousands of lipids. The HAS approach was demonstrated for a model with one bead per lipid (Figure 2c) and the HCG with 3–4 beads per lipid. The neat feature of these models is that analytical potentials describing the generic behavior of the lipids are combined with detailed FM potentials that give the model chemical specificity.

CG Protein Models

CG models for proteins have a long history, with the pioneering models for protein folding introduced in the mid 70s[75] and 80s.[76] The motivation for such simple models was, as for most biological molecules, to address the issue of conformational sampling. Structure-based models have contributed to our understanding of the physicochemical forces governing the protein folding process[77] and protein–protein interactions.[78] However, these models often lack a proper description of the chemical specificity of amino-acid side chains (SCs) and are therefore not described here. Readers are referred to earlier reviews on such models.[79,80] CG protein models that retain chemical specificity are diverse with respect to the level of representation (Figure 3) and complexity of the associated interaction potentials, which are closely tied to the problem of interest. In particular, a detailed backbone (BB) is compulsory for exploring secondary structure formation while SCs are more important for protein–protein interactions. These two sets of interactions may thus be parameterized separately and may differ in the level of mapping, varying from one to five CG sites per residue for both the BB and the SCs. Most protein CG models use a combination of physics-based and knowledge-based potentials for which transferability is probably the most challenging aspect. Some models have been quite successful and contributed to the popularization of CG approaches as an alternative to atomistic models.

Bereau and Deserno Model

The protein CG FF developed by Bereau and Deserno[81] uses an intermediate level of description emphasizing on structure. The model has a quite detailed three-bead protein BB and one-bead SCs (Figure 3a). The bonded terms were derived from existing geometric parameters and given an approximate 5% flexibility around their reference values to account for thermal fluctuations. The BB phi/psi dihedral angles were used as parameters during the fitting and as an indicator of local structure and flexibility. The SC nonbonded interactions were based on the knowledge-based potential derived by Miyazawa and Jernigan[85] from a statistical analysis of SC contacts in a protein structure database. Bereau and Deserno converted this energy scale into a two-body distance-dependent potential. Note that the model does not account explicitly for electrostatic interactions but they are implicitly accounted for in several terms. The BB beads interact through a more complex combination of terms that have previously been identified as secondary structure determinants. These include local excluded volume, an explicit geometric H-bonding function, and a dipole–dipole interaction of neighboring residues. These terms were tuned to reproduce local protein structure (Ramachandran plot or dihedral BB distribution of GGG and GAG tripeptides) and global folding properties (folding of a three helix bundle). The FF showed promising results in folding α-helical proteins[86] but fine-tuning is needed to stabilize β-sheet structures and proteins with a mixture of α-helices and β-sheets.

The OPEP Model

The OPEP (Optimized Potential for Efficient Protein structure prediction) CG model developed by Derreumaux and coworkers[82,83] is a generic CG model. It has a detailed BB close to a full atomistic model (N, HN, Cα, C, and O atoms are represented) and uses a single CG bead for each SC, with the exception of the proline SC, which has three beads (Figure 3b). The position of the SC is defined by the BB conformation using an off-lattice (discrete) representation.

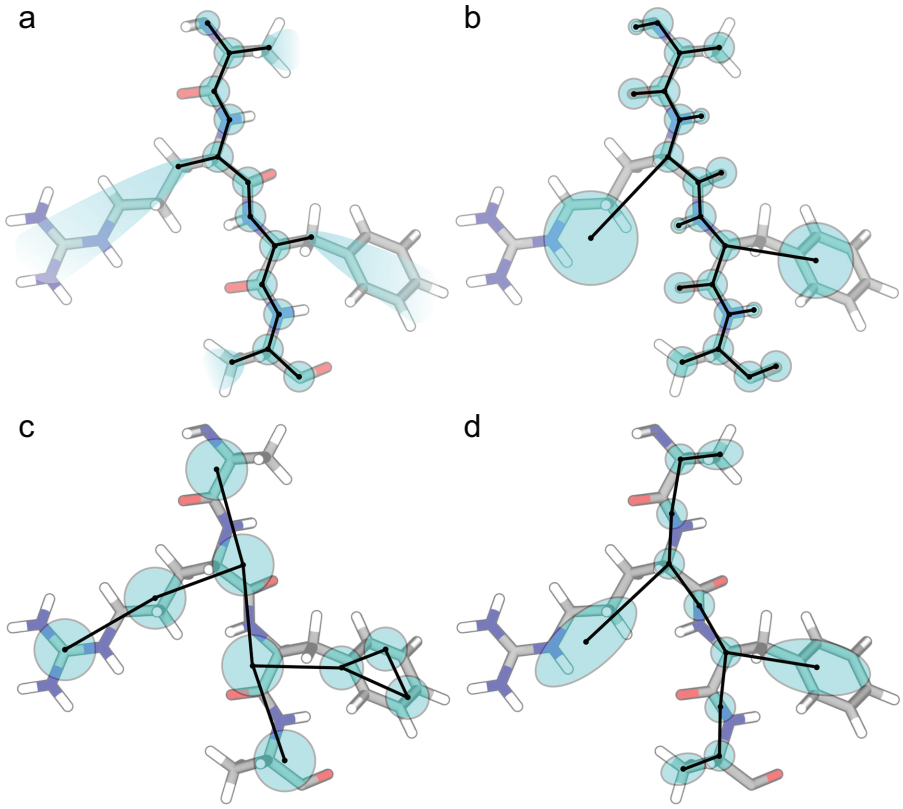

FIGURE 3 | Mapping strategies for CG protein models illustrated for an AlaArgPheAla peptide. (a) In the model from Bereau and Deserno,[81] the CG particle for the side chain is located on Cβ but the effective vdW radius is for the entire side chain. (b) In the OPEP[82,83] model the backbone H-atom are represented explicitly. (c) In the Martini[84] model the backbone bead is at the center of mass of the non-H atoms and its type is secondary structure specific. (d) In the UNRES[24] model the Cα is a virtual site, the interaction sites are the peptide-bond and the side chain ellipsoids.

The most recent OPEP potential (version 4.0) is a combination of generic-bonded terms (derived from the all-atom AMBER FF) and nonbonded interactions consisting of vdW and H-bonding terms. The vdW interactions are knowledge based and combine BB–BB, BB–SC terms with 210 SC pair interactions. The H-bond potential combines a two-body geometry-dependent term with a four-body term to account for H-bond cooperativity. The effects of solvent are taken into account implicitly within the nonbonded terms. The OPEP potential was parameterized to maximize the energy gap between native and nonnative structures and to enforce stability of native structures in MD simulations. OPEP was used successfully for protein folding,[87] structure prediction[88] and aggregation studies.[89,90] Potential drawbacks of OPEP models are the lack of SC specificity crucial for accurate description of protein–protein interactions and the need of a 1.5 femtoseconds integration time (due to the detailed BB and H-bonds) in MD simulations, which reduces the amount of conformational sampling as compared with other CG models.

The Martini Model

As an extension to the Martini lipid FF (see above), the protein version[84] has inherited the general 4:1 mapping used to define chemical groups as beads and the broad experience in using partitioning data for parameterization. Each amino acid is represented by one bead for the BB and from zero (Gly and Ala) to four (Trp) beads for the SC (Figure 3c). The bonded terms were extracted from a set of protein structures with the BB bead placed on the center of mass (COM) of the BB[84]. An elastic network model (based on the Cα positions) was parameterized in conjunction with Martini to improve structural stability.[91] Partitioning behavior of SC analogues between water and oil phases and at the water interface of a DOPC bilayer was originally used to parameterize the nonbonded interactions. Recently, a thorough examination of the binding of Wimley–White pentapeptide to a palmitoyloleoylphosphatidylcholine bilayer[92] and SC analogue (self and cross) dimerization free energies[93] were used to refine the nonbonded parameters.[94] Notably, an explicit polarization term was added

to polar SCs such as glutamine and asparagine. Coulomb interactions for charged SCs use the standard shift function used in Martini with a relative dielectric constant $\varepsilon = 15$ or $\varepsilon = 2.5$ when combined with the regular and polarizable water, respectively. The Martini FF does not allow dynamic secondary structure conformational flexibility and thus precludes folding studies. However, it has been successful in describing protein tertiary conformational changes,[95] protein supramolecular organization,[96–98] and their relation with the membrane environment.[99,100]

The UNRES Model

The UNRES (UNited RESidue) CG model developed by Liwo et al.[24] models the BB by two CG beads—an interacting peptide-group (P) and noninteracting (CA) group—and the SC as a single ellipsoidal bead (Figure 3d). This FF has gone through numerous refinements[101] to come close to a pure physics-based version in contrast with its first appearance two decades ago as a strongly knowledge-based potential.[102] The UNRES potential is a free energy function where all the DOFs (including that of the solvent) are averaged out into effective potentials, except for the ones describing the protein conformation. The bonded interactions include bonds, angles, and dihedrals for the BB and a rotational potential defines the SC rotamer. The nonbonded interactions (vdW and Coulomb) include terms between SC–SC, SC–P, and P–P beads. All the nonbonded terms are derived from *ab initio* or semi-empirical calculations of small model systems and PMFs extracted from all-atom MD simulations of SC analogue pairs. Notably, the UNRES potential incorporates temperature-dependent correlation terms. UNRES now stands as a prototype for purely physics-based approaches to coarse graining and it has been successful and widely used over the last two decades to study protein folding,[103] structure prediction,[104] protein–protein binding,[105] and mechanisms of protein fibrillation.[106] However, there are a couple of caveats. First, in contrast to other CG models, the conversion from UNRES to an all-atom representation is not straightforward and may thus be inappropriate for multiscale approaches. Second, because of its emphasis on accurate description of protein interactions, it will be difficult to extend UNRES to other biomolecules such as nucleic acids and lipids.

The SCORPION Model

The SCORPION protein CG FF[107] was initially developed as a scoring function for protein–protein recognition using one bead to model the BB and one or two beads for the SCs. The bead self- and cross-interactions were extracted by (1) fitting vdW SC pair interactions extracted from PMFs of SC association determined at an atomistic resolution and vanishing charges and (2) determining a set of point charges to reproduce the electrostatic potential, the total charge, and the permanent dipole as described by an atomistic model of the full protein. The parameterization was done in vacuum to allow, in principle, mixing with any solvent. In a recent study,[108] the authors combined the protein potential with a compatible water model, which was validated against solvation free energies of peptides. The use of the combined protein and water model showed great promise for studying protein–protein recognition.[108] The notorious and challenging barnase/barstar complex and two others were successfully predicted. The current main drawbacks of the model are the lack of bonded terms (an elastic network is used instead) and the use of a high temperature in the simulations, which affects the balance between the enthalpic and entropic contributions.

Other Promising Models

The PaLaCe model recently introduced by Lavery and coworkers[109] to study the mechanics of proteins uses a two-tier representation of amino acids: one for bonded and another one for nonbonded interactions. The FF consists of physics-based bonded and nonbonded terms combined with an implicit treatment of the solvent and a BB H-bonding potential (allowing secondary structures changes). These terms were collectively and iteratively parameterized against a large database of protein structures and MD simulations of their native state. The PaLaCe model was able to reproduce force-induced conformational changes for the immunoglobulin-like domain of the giant protein titin, originally observed by single-molecule experimental results and all-atom simulations.

The PRIMO model proposed by Feig and coworkers[110] represents the protein BB using three CG beads and one to five sites for the SCs. The mapping was carefully optimized to allow a high-resolution reconstruction of all-atom protein models[111] aiming for multiscale approaches. The interaction scheme is typical of a molecular mechanics potential with bonded and nonbonded terms optimized against a diverse set of peptides and proteins described by the all-atom CHARMM FF. The model also includes an explicit treatment of BB H-bonds and an implicit solvent. The PRIMO FF was validated against the conformational sampling of alanine-based polypeptides and folding of small

peptides as observed in atomistic MD simulations and experimental data.

In the AWSEM (Associative memory, Water mediated, Structure, and Energy Model) FF, developed by Papoian and coworkers,[112] the position and orientation of each amino acid residue is dictated by the positions of its Cα, Cβ, and O atoms. AWSEM combines a large number of physical interactions, from BB terms to direct and water-mediated interactions and H-bonding, with structural biases that are local in sequence, based on the alignments of fragments of nine residues or less of the target protein to the local segments found in a protein database. It has been successfully used to predict protein structures both *de novo* and using homology models,[112] as well as dimeric protein interfaces.[113] The dynamic properties of the model have yet to be characterized in detail.

CG Nucleic Acid Models

Nucleic acids (DNA and RNA) seem more tractable for modeling than proteins due to the smaller number of building blocks involved. Nevertheless, perhaps because of scarcer structural data, tight packing, and higher charge density, the development of CG nucleic acid models has progressed more slowly than with many other biomolecules.

Despite the similarities between DNA and RNA, the challenges in modeling them differ greatly. DNA exists primarily as double-stranded structures (dsDNA) with only a limited number of well-defined conformations but forms extremely large-scale assemblies. For RNA, the challenge is to predict how the single-stranded RNAs (ssRNAs) fold into their functional form. These differences have led the CG models for DNA and RNA to use disparate strategies in order to reach their objectives. Most CG RNA models are aimed at predicting structures (see a recent review[114]) and use structure-based potentials. Only a few CG RNA models can be used in MD simulations,[115,116] in contrast to the numerous DNA CG models available—which span several orders of magnitude in the length scales they describe. Large-scale mechanical properties of DNA structures have been studied using very coarse models since the early 90s.[117,118] The development of more detailed CG models that are able to describe for example DNA melting started later.[119] We focus here on CG models that are detailed enough to describe sequence specificity. For an overview of coarser models, we refer the reader to previous reviews.[120–122]

3SPN Models

de Pablo and coworkers[27] proposed a CG model for DNA, coined 3SPN.0, where phosphate, sugar, and base are described using one bead each (Figure 4a). Bonded interactions were derived from a canonical structure of B-DNA. Base stacking was implemented using intrastrand Gō-potentials that act between the first and second neighbors. H-bonds were modeled using a potential that acts between complementary bases and an excluded volume term is used between all other beads. Electrostatics is described using a Debye–Hückel approximation. The authors did note that these choices bias the model toward the B-form of DNA compared with other forms. The parameters were selected to reproduce experimental melting curves of DNA at a specific salt concentration, but the model also performed rather well in other salt concentrations as well as reproducing the persistence length of dsDNA and ssDNA and bubble formation in dsDNA. The authors measured roughly three orders of magnitude speedup as compared with atomistic simulations for short dsDNAs. The model has been refined (3SPN.1)[123] to describe DNA hybridization and to model ions explicitly.[124–126] Very recently, the 3SPN.2[127] model that does not use Gō-potentials for stacking was introduced.

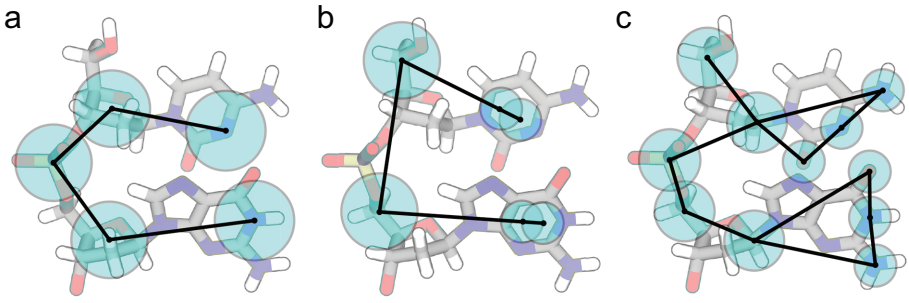

FIGURE 4 | Mapping strategies for CG nucleotide models illustrated for cytosine (top) and guanine (bottom) based on (a) 3SPN.0,[27] (b) Ouldridge,[25] and (c) Dans[28] models.

Ouldridge Model

Ouldridge et al.[25,128] developed a model that represents the nucleotide as a rigid system of one BB site aligned with two sites for the base (Figure 4b). Neighboring nucleotides interact with excluded volume and angle-dependent stacking potentials, whereas the BB beads are connected with finite extensible nonlinear elastic bonds. All other beads interact with excluded volume, cross-stacking, and H-bonding potentials. The latter two are also angle dependent. All potentials except for H-bonding are identical for different nucleotide pairs. The model does not include electrostatics but it is parameterized in high salt concentration where they play a smaller role. Furthermore, solvent effects are included implicitly within the interaction potentials. The potential functions are parameterized to reproduce experimental properties of base stacking in ssDNA and dsDNA melting. The parameterized model was shown to reproduce mechanical properties of both ssDNA and dsDNA as well as DNA hairpin formation. The stacking and H-bonding interactions were further refined[129] to introduce sequence dependence and reproduce experimental melting temperatures of short duplexes. The model was also able to show sequence-dependent differences in the structure of ssDNA as well as in the opening of dsDNA ends.

Dans Model

Dans et al.[28] developed a DNA model for the SIRAH FF that describes each nucleotide with two interaction sites for the BB, one for the sugar, and three for the base (Figure 4c). This mapping allows for easy backmapping to atomistic detail. The bonded interactions were parameterized to reproduce the canonical B-DNA structure. The nonbonded interactions are described using LJ and Coulomb potentials and were parameterized to reproduce structural and dynamical properties of dsDNA as observed in atomistic simulations. A generalized Born model describes the solvent. The authors found qualitative agreement with experimental melting curves of dsDNA, dsDNA transition from A to B DNA structures and base pair opening dynamics as observed in long atomistic simulations. The model was roughly 600 times faster than fully atomistic simulations for short dsDNAs. The model was later adjusted to include the explicit solvent model WT4.[22] The modified model reproduced qualitatively effects such as experimentally observed narrowing of the minor groove due to cations, but the increased detail rendered it roughly 30% slower than the implicit model.

Other Promising Models

The recent HiRE-RNA model by Pasquali and Derreumaux[130] is a physics-based model for RNA folding, similar to the OPEP model for proteins. The model uses one bead for the phosphate, four beads for the sugar, and one or two beads for the base. It employs standard bonded potentials, OPEP type LJ potentials, and explicit H-bonding potentials. It does not describe ions or solvent and includes electrostatics only between phosphates. The current model is an unoptimized proof-of-concept model but shows promising results in folding small RNAs using replica exchange MD simulations.

A DNA model by Linak et al.[131] has the same mapping as the 3SPN models but uses directional nonbonded potentials to describe both the Watson-Crick and Hoogsteen base pairing and employs no dihedral potentials. The model shows improved accuracy in situations where Hoogsteen pairs are known to play a role but has difficulties with the handedness of dsDNA due to the omitted dihedrals.

The model by Morriss-Andrews et al.[132] describes each nucleotide with three beads and uses orientation-dependent base–base and base–BB potentials as well as an explicit hydrogen bond potential. Bonded and nonbonded parameters are derived from atomistic potentials and distributions from atomistic simulations. The model reproduces the structure and chirality of dsDNA as well as persistence length of both ssDNA and dsDNA.

CG Carbohydrate Models

Carbohydrates are ubiquitous biomolecules involved in many biological processes. Because of their nature, carbohydrates encompass a huge degree of polymerization making up a virtually infinite number of sequences, linkages, and degrees of branching.[133] Unlike proteins, nucleic acids, and lipids, which tend to predominantly adopt a relative well-defined (native) conformation under the conditions where they are biologically functional, carbohydrates typically feature a high degree of conformational freedom.[134–136] As a result of this structural diversity, carbohydrates represent a very challenging class of biomolecules in terms of CG modeling. Consequently, current CG carbohydrate models often aim for the simulation of very specific systems.

The MB3 Model

Pioneering efforts by Molinero and Goddard[39] aimed at modeling hexopyranose glucose. This model, coined 'MB3', was the first attempt to build a

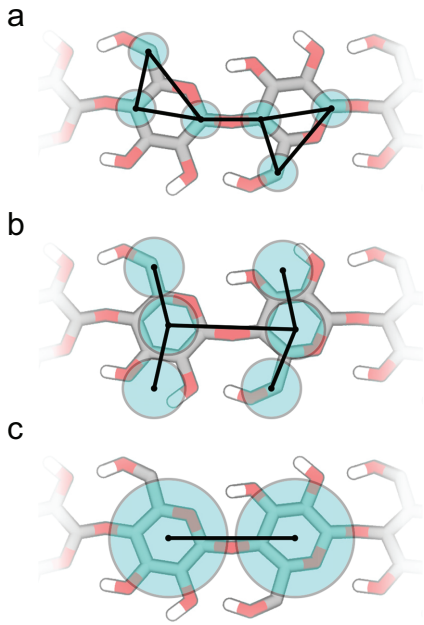

FIGURE 5 | Mapping strategies for CG carbohydrate models illustrated for a single cellulose fibril. (a) The MB3[39] model is based on the atomistic position of C1, C4, and C6 for every hexopyranose. (b) The Martini[138] model relies on the COM of the atoms enclosed by the circles. (c) The solvent-free model of Srinivas et al.[26] makes use of a single bead for the representation of every monosaccharide subunit.

robust reductive model for the simulation of carbohydrates. The hexopyranose ring is represented by three particles (Figure 5a) and was directly mapped from atomistic simulations, with bonded terms including bond, angle, and dihedral potentials derived using an IBI approach. A single nonbonded term was used for all interactions and rigorously parameterized against density, cohesive energy and structural unit cell parameters. While the model was successfully applied to the simulation of disaccharides (maltose) and longer polysaccharides (amylose), the set of bonded parameters is state specific and thus not transferable to alternative glycosidic links. Following the same, topological descriptors, Liu et al.[137] developed a reductive model for the simulation of carbohydrates, but using nonbonded terms based on more versatile pretabulated potentials derived from atomistic simulations. This model can be combined with an explicit water model. So far, applications of the model have been restricted to the simulation of glucose and amylose.

The Martini Model

A more extendable and general approach is the one taken by López et al.[138] based on the Martini CG FF. Each saccharide is mapped using three CG beads in such a manner that the underlying polar–nonpolar feature of the ring is preserved (Figure 5b). Following the philosophy of the Martini model,[43] several monosaccharides and disaccharides were parameterized combining bottom-up and top-down approaches. Atomistic trajectories were used to iteratively adjust the set of bond, angle, and dihedral potentials of the CG representation. Nonbonded terms were determined to reproduce experimental partitioning data. While the parameters are straightforwardly applicable for the simulation of mono-, di-, and polysaccharides in the colloidal state, application to the crystalline phase is rather problematic as was shown by Wohlert and Berglund[139] in the case of crystalline cellulose. A potential advantage of the Martini carbohydrate model is its compatibility with the ample set of different biomolecules, illustrated recently by the parameterization of a Martini FF for glycolipids[140] and application to cyclodextrin–cholesterol complex formation.[141]

Bellesia Model

Recently, Bellesia et al.[142] developed a solvent-free CG model for the interconversion between cellulose I_β to III. Based on a five-bead mapping of the ring, the model combines LJ terms with harmonic bonded potentials aimed to reproduce the crystalline phase of cellulose. The CG representation not only reproduces the torsional angles between glucose planes, but also the transitional rotameric states of the hydroxymethyl groups thus effectively mimicking the changes in both intracrystalline hydrogen bonds and stacking interactions during the transition from cellulose I_β to cellulose III. The model has been shown to reproduce structural as well as thermomechanical properties of cellulose.

Bathe Model

The solvent-free model of Bathe et al.[143] is aimed at modeling chondroitin (glycan chains forming a major component of the extracellular matrix). This model explicitly describes the DOFs associated with the torsional angle representing the glycosidic linkage. Each monosaccharide consists of three-bead sites, plus two additional beads at the centers of charge and geometry, used to model the nonbonded electrostatic and steric interactions, respectively. All-atom resolution trajectories of isolated disaccharides are used to generate pretabulated PMFs for the glycosidic torsions. Electrostatic interactions are included between nonadjacent monosaccharides using a Debye–Hückel potential, assuming zero ionic radius. Steric interactions between nonadjacent monosaccharides were modeled using a LJ potential applied to the center of geometry. The model was able to reproduce

the ionic strength-dependent persistent length, pH-dependent expansion factor, and titration behavior of chondroitin.

Srinivas Model

Aimed to study the structure and dynamics of cellulosic biomass, the model developed by Srinivas et al.[26] pursues to study the intrinsic conformation of long I_β cellulose macrofibrils. Their simplified representation makes use of a single bead for every monomeric glucose subunit (Figure 5c). The glucose COM was mapped from atomistic simulation trajectories and every associated bond, angle, and torsion potential parameter was extracted from the same conformational ensemble. Nonbonded interactions were optimized using an IBI approach, with the distance distributions between individual monomers as target observables. The model has been used for studying the transition between crystalline and amorphous phase at long time scales, as detailed in the *Applications of CG Models* section.

Other Promising Models

Recently, Satelle et al.[144] developed an interesting CG model for the prediction of hydrodynamic properties of heparin sulfate. The CG model potentials were carefully adjusted to reproduce the glycosidic linkage between consecutive monosaccharide subunits and the internal ring puckering observed in long unbiased all-atom simulation. The model is not only able to reproduce relative ring–ring orientations but also the internal energy landscape of different ring conformers.

Markutsya et al.[145] constructed four different CG models of cellulose, with potentials derived from FM either using one, two, three, or four sites per monomer. They found that the four-site CG model is most promising, as it is best at reproducing the glucose–glucose conformations observed in the all-atom simulation. The model underscores the importance of decoupling the pyranose ring from the oxygen atom in the glycosidic bond when developing all-atom to CG mapping schemes for polysaccharides.

APPLICATIONS OF CG MODELS

Applications of CG models range far and wide and providing an exhaustive coverage is far beyond the scope of this review. Instead we cherry-picked five state-of-the-art examples demonstrating successful and potentially inspiring use of CG models. We show different types of application with various aims and include a variety of biomolecule classes as discussed in the *CG Models* section. We start by the description of typical applications of a protein CG FF whose development emphasized on getting the native protein structure, following by larger scales applications using a model which allows for exploration of protein conformational changes and assembly in lipid bilayers, ending with coarser models of DNA hybridization and cellulose fibrils stability.

Protein Folding

In spite of a few recent studies able to simulate the folding of a number of small proteins,[146–148] the current state of atomistic protein folding simulations is mostly limited to small single-chain proteins using considerable dedicated computer resources. CG models are thus extremely attractive but capturing the relevant DOFs has shown to be quite challenging.[122] In that context, the UNRES[24,101] model is unique. Its development over more than two decades has gone through many successive but consistent modifications. This laborious parameterization illustrates the extreme challenging aspect of the development of a reliable CG FF when it is about capturing a delicate mixture of complex structural and chemical contributions. UNRES successfully participated in the CASP exercise in which success of *ab initio* prediction of new protein folds is almost exclusively reserved to knowledge-based potentials. Two UNRES predictions are shown in Figure 6a: the target T0215 (a three-helix bundle; PDBX9B) and T0281 (a α/β fold; PDB:1WHZ) predicted to 0.35 and 0.55 nm Cα RMSD from the native state.[104] The latter represents one of the first successful predictions of an α/β structure with a physics-based potential. UNRES also described the folding pathway of several single- and multichain proteins. For example, UNRES folded *ab initio* (Figure 6b) the 48-residue Lysm domain protein (PDB:1E0G) to a structure with a 0.39 nm Cα RMSD from the native structure.[149] The folding processed by the initial formation of an almost all α-helical structure followed by the unfolding and refolding of the C- and N- terminal regions into their native antiparallel β-sheet structure. Another example of *ab initio* folding with UNRES is of a multichain protein (PDB:1G6U) (Figure 6c). Individual chains folded independently to their native structures and later assembled into a structure at 0.24 nm Cα RMSD from the native state.[150]

Gating of Mechanosensitive Channels

The bacterial mechanosensitive channels of large conductance (MscL) serve as a last resort emergency

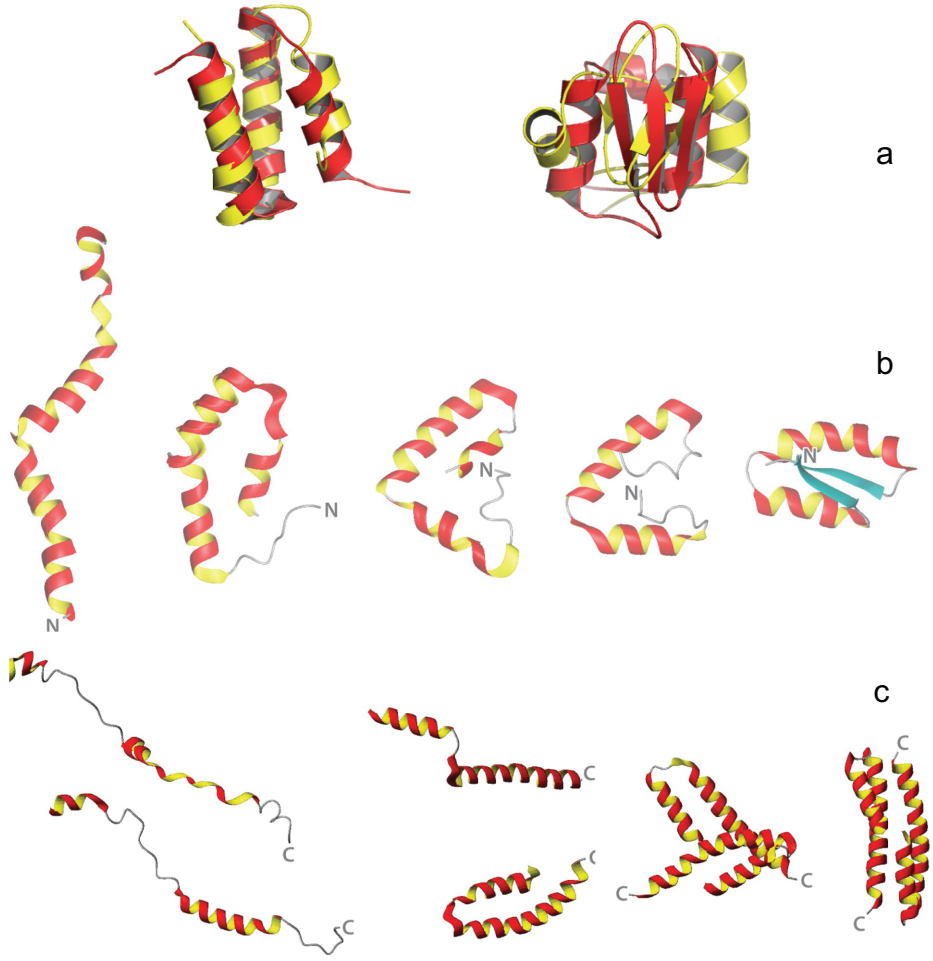

FIGURE 6 | CG protein folding. (a) Predicted structures from the UNRES[24] CG FF in CASP6 exercise for targets T0215 (left) and T0281 (right). The native structure is colored red and the predicted structure yellow. (b) Snapshots from an *ab initio* folding of a 48-residue Lysm domain protein and (c) a synthetic domain-swapped protein consisting of two 48-residue chains. Figure 6b reproduced with permission from Ref 149. Copyright 2005, PNAS. Figure 6c reproduced with permission from Ref 150. Copyright 2007, American Chemical Society.

release valve, protecting bacteria from lysis upon acute osmotic downshock. With induced membrane tension, MscL opens a large (~3 nS) mostly unselective pore, releasing ions and small solutes, thereby relieving the cytoplasm of osmotic tension.[151] Because of the high computational cost of atomistic simulations, MscL gating from its closed state[152] has not been simulated at the atomistic level without strong biasing potentials.[153–155] It was only recently that MscL could be gated in an unbiased way and in tractable computational time using the Martini CG FF.[95,156,157] Those studies confirmed the iris-like opening mechanism of the channel and provided valuable insight into how changes in protein shape influence its preferred conformational equilibrium. To gate MscL in CG simulations, the channel was equilibrated in a solvent/bilayer environment for a few microseconds after which tension was rapidly applied. In 10–100 nanoseconds following the application of lateral tension and thinning of the bilayer, the MscL transmembrane helixes tilted, extending the extracellular cavity of the channel. The channel hydrophobic gate takes an additional 0.2–2 microseconds before rapidly expanding and opening the channel (Figure 7). Such gating of MscL took 5–10 days on a 12-CPU computer at the CG level; it is not clear how many years it would take at the atomistic level, even using much faster computers. Ongoing work on MscL has shed new light on its gating mechanism and how changes in bilayer properties can influence membrane protein gating. Moreover, this CG model of MscL provides a new tool for studying the mechanism of mechanosensation, the bilayer regulation of membrane proteins, and the rational design of future drug delivery systems, for example, Ref 158.

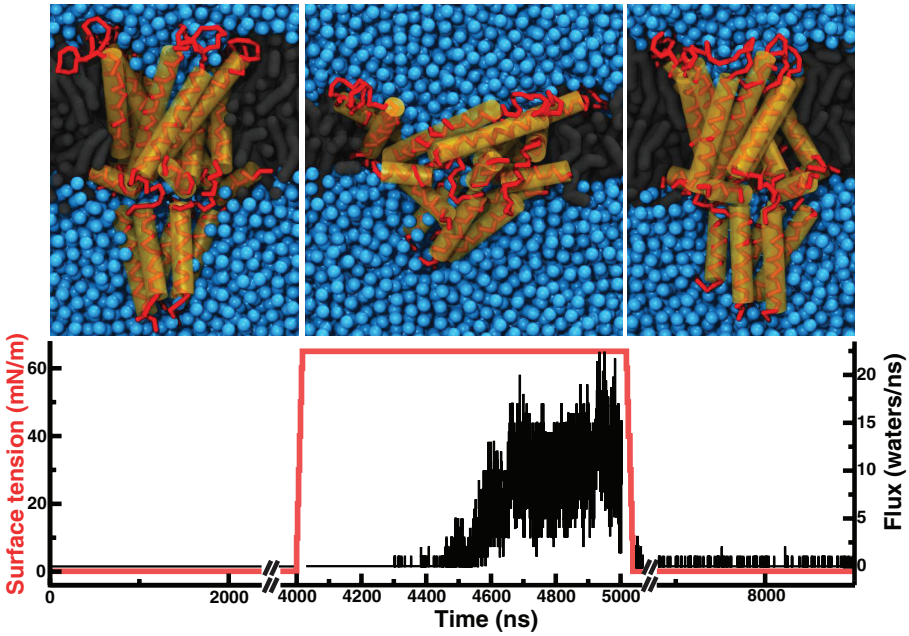

FIGURE 7 | Reversible gating of MscL using the Martini CG model. A MscL is solvated in a DOPC bilayer and equilibrated for 4 μs (left). Surface tension is then applied to the bilayer, the channel opens and CG water permeates through the large channel opening (middle). When surface tension is removed the channel recloses and water flux is almost completely abolished (right).

Membrane Protein Self-Assembly

Biological membranes have a complex and dynamic supramolecular organization that has recently emerged as a potential major component in many fundamental processes. The transient nature of these processes has made their characterization by conventional experimental and computational approaches a great challenge. CG MD simulations using the Martini model have shown promise in elucidating the forces involved at the molecular level with a close to atomistic resolution. Notably, a set of recent studies took advantage of the increased system size and length scales accessible to reveal a few significant protein/lipid interplays. First, a lipid membrane responds to the presence of a protein by an anisotropic deformation at the protein/bilayer interface in order to match to the protein's hydrophobic surface.[96] Second, the extent of membrane deformation determines the protein's propensity to self-organize.[96] Third, the protein surface properties (sequence dependent) design specific lipid binding sites[100] and favor different protein/protein interfaces that may drive proteins to assemble into well-ordered and highly organized arrays.[97] Forth, protein sorting is mediated by lipid properties in multidomain membrane patches.[159] A typical system used in these studies is pictured in Figure 8. It consists of 64 visual receptor rhodopsins embedded in a DOPC membrane bilayer at 1/100 protein/lipid molar ratio.[97] The 100 microseconds time scale reached in these simulations will soon become a standard and the millisecond time scale is not far off.

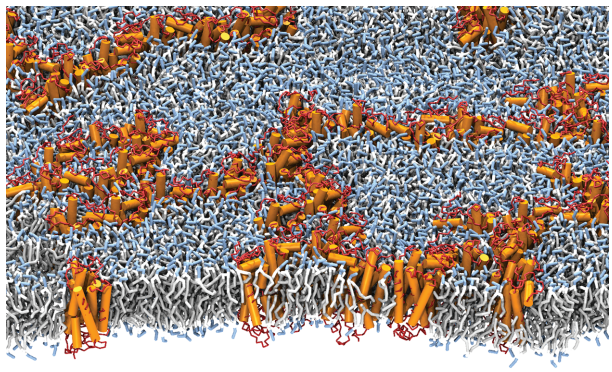

FIGURE 8 | Membrane protein self-assembly. Snapshot of 64 visual receptor rhodopsins in a DOPC bilayer. The receptors were initially placed on a 8 $\times$ 8 grid and left free to self-assemble for a period of 100 μs. The receptor's transmembrane helices are shown as orange tube and the backbone trace in brown. The lipid head groups are shown in light blue, the glycerol moieties in white, and the tails in gray.

DNA Hybridization

DNA hybridization is the assembly process of complementary ssDNAs forming dsDNA that is ubiquitous

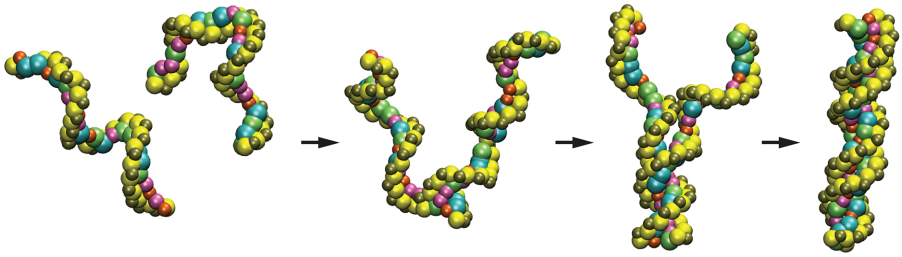

FIGURE 9 | DNA hybridization process. Starting with two separate strands, they first associate and form a nucleation site.[127] Bases make complementary pairs leading to a fully hybridized dsDNA. The data for the figures are produced with the recent 3SPN.2 model. Figure kindly provided by Dan Hinckley and Juan de Pablo.

in, for example, transcription. Also a large number of systems and protocols in molecular biology and biotechnology rely on DNA hybridization. Therefore, the understanding of the sensitivity of this process to the sequence of the ssDNA strands, as well as to the effect of environmental factors like temperature or the salt concentration, is essential. MD simulations offer ways to bridge the gap between experimental observations and theoretical models and elucidate the whole hybridization process by describing the hybridization pathways and kinetics. Such studies are particularly amenable for CG models that are able to reach system size and time scales relevant to study such pathways thoroughly. Sambriski et al.[160] used the 3SPN.1 CG DNA model[123] coupled with a transition path sampling technique to investigate the hybridization process of 14–30 base pairs long ssDNAs. These simulations revealed the presence of multiple and nonspecific hybridization pathways for highly repetitive sequences, whereas for random sequences, a nucleation site of specific complementary base pairs is formed to start the hybridization process (Figure 9). In the latter case, the pathway from initial nucleation is strongly restricted. In addition, they could pinpoint short repetitive sequences as the most probable nucleation sites because of greater number of possible complementary base pairs. This work illustrates how CG simulations can be successfully applied to understand the mechanisms underlying the experimental observations of a crucial cellular process.

Cellulose Fibrils

Considerable attention has been focused on the study of plant biomass, especially concerning the degradation of cellulose and its application to biofuels. In that respect, computational modeling has been applied widely to understand the structure and conformation of cellulose. Natural cellulose microfibrils are on average several micrometers in length. The large

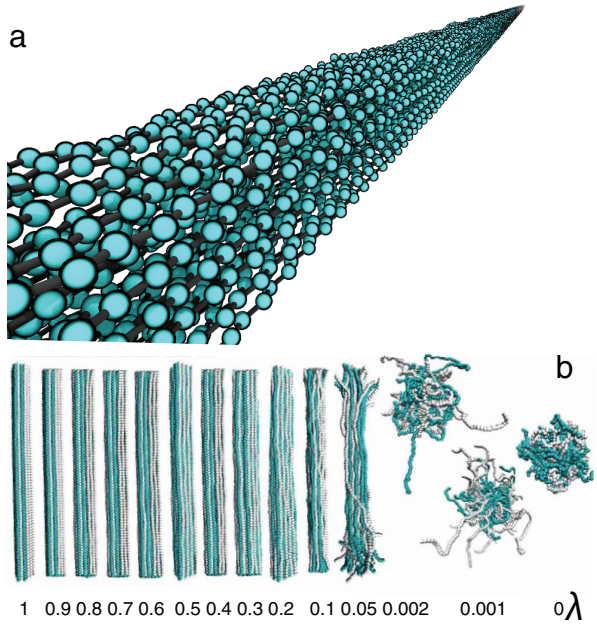

FIGURE 10 | Cellulose microfibril. (a) Representation of the crystalline I_β cellulose microfibril according to Srinivas et al.[26] model. (b) Transition of cellulose from its crystalline state to the amorphous conformation. The structural change is dependent on the tuneable LJ coupling factor λ. Reproduced with permission from Ref 26. Copyright 2011, American Chemical Society.

degree of polymerization makes it virtually impossible to obtain reliable conformational data using atomistic models. Using a CG model, Srinivas et al.[26] studied the structural differences between cellulose fibrils and amorphous cellulose. By tuning of the LJ potential through a λ factor, a discrete transition between the fully crystalline ($\lambda = 1$) to the fully amorphous ($\lambda = 0$) systems was established. In line with experimental evidence, the model suggests that during the conformational transition, the microfibril denaturation is started by an external uncoating mechanism that involves the outer cellulose fibrils (Figure 10). Because of the versatility of the model, different cellulose

crystal allomorphs can be studied and characterized at permissible computational effort.

OUTLOOK

Why are CG Models Useful?

During the last decade, we have seen a thriving development of a large number of CG biomolecular models, of which we have only been able to provide a limited outline above. However, do we actually need any of these CG models? This may seem like a provocative question, but considering the continuous increase in availability of computational power (soon entering the exascale era), one may argue that most relevant (bio)materials can be studied at the atomistic level in the near future. Currently, the largest systems that can be handled by particle-based simulations are limited to 10^7 interacting atoms and time scales up to 1 microsecond, but we can expect these limits to steadily increase (according to Moore's law, computational performance doubles every 2 years). Could we thus soon model a complete cell in atomistic resolution? Not really. Modeling a typical eukaryotic cell, for instance, with a diameter of 10 μm requires about 10^{14} atoms. Given the relevant time scales on which cellular processes take place, microseconds to seconds and beyond, it is clear that simulations capturing the full complexity of a cell in atomistic detail are still far fetched. Even the kind of CG models described in this review will have a hard time to cope with this size, but CG modeling of a bacterial cell with a much smaller diameter of approximately 0.5 μm, amounting to 10^9 atoms, becomes tractable in the foreseeable future. Obviously, many cellular processes can be studied at smaller length and time scales, but even a simple case like simulating the undulation spectrum of a planar lipid bilayer requires CG modeling as soon as one increases the system size above 100 nm. This example was presented in an eloquent essay by Deserno,[161] in which he argued that the computational effort needed to increase the system size of a typical membrane patch of length L scales as L^6 due to the much longer relaxation times involved. Thus, increasing the typical size of an atomistic membrane patch (around 20 nm) to 200 nm would require a million times more work to fully equilibrate the system. Even assuming Moore's law continues to proceed apace, we would still need to wait for about 40 years to accomplish this! Thus, to simulate collective effects such as the formation of multimeric membrane protein complexes, membrane patches of 100 nm size and beyond are necessary, which in the foreseeable future can only be accomplished at the CG level.

Current Challenges of CG Model Development

The current generation of CG models are still under active development and numerous new models are continuously emerging. One may wonder, is there not an optimal CG FF that eventually may replace all current models? The answer is no. At a given state point, the pair potential that reproduces the pair structure is unique (see Box 1), implying that it is impossible to simultaneously represent the pair structure and additional key thermodynamic properties of the system with pair potentials. This is known as the representability problem.[33] In practice, this means that a different model is required depending on the question asked. To choose the right model for each problem, it is necessary to know the limitations of the model. Therefore, in this review, we have tried to indicate the underlying assumptions of commonly used CG models. However, a number of limitations are pertinent to most CG models and future efforts should be directed to improve on those aspects: (1) the model is too biased, that is, not transferable to other state conditions, (2) the model is only parameterized for a specific class of molecules, implying there is a lack of compatibility, and (3) the model is too coarse to capture certain behavior.

To improve transferability, systematic frameworks for obtaining accurate CG potentials from higher resolution data are being developed. For instance, CG potentials can be simultaneously optimized with respect to multiple reference simulations[162] and a variety of experimental data can be used as additional constraints in the optimization strategy,[163] essentially combining top-down and bottom-up approaches. Automated workflows are being developed to generate large sets of converged interaction potentials.[163,164] Methods have been developed to measure and minimize the information loss upon coarse graining[14] and to optimize the mapping procedure.[165] Additionally, ongoing improvements of atomistic FFs provide us with more accurate reference structures and the steady increase in single molecule experimental data (e.g., force spectroscopy of individual biopolymers, single molecule particle tracking) allow for novel and direct ways to further calibrate and validate our models.

The issue of limited compatibility is related to the lack of transferability (and to the general representability problem), but there is a pressing need for compatible FFs that can be used for more than just a single class of biomolecules. Except for the generic Martini model, none of the CG FFs discussed above can describe the more complex setting of real

biosystems, let alone handle the rich and growing diversity of bioinspired materials such as biofunctionalized nanoparticles, DNA–polymer hybrids, peptidosurfactants, and so on. Triggered by the development of more transferable models, we expect an increase in the number of compatible CG FFs in the near future.

The problem of a model being too coarse is touching upon the very limits of coarse graining. No matter how hard we try, not every problem can be tackled with a CG model. Some atomistic details are notoriously hard to mimic at a CG level, for example, the directionality of H-bonding, and sometimes fine grain resolution is required. In that respect, the active field of multiscaling[166–171] shows a lot of promise. Multiscale methods treat part of the system, the region of interest, at high resolution and the surrounding at a lower level of resolution, thereby combining the advantages of atomistic and CG models. Multiscale methods can either use a static division as in QM/MM, or allow particles to change resolution on the fly. The challenge is to achieve a realistic coupling between the atomistic and CG DOFs. One route is to specifically parameterize the cross-interactions, as demonstrated in a number of recent test systems.[62,172,173] Although the results are encouraging, such methods are not easily transferable. A more generic approach is the AdResS (Adaptive Resolution Simulation) scheme, developed by Kremer and coworkers.[174] In this method, a transition region allows molecules to pass from atomistic to CG resolution and vice versa as a function of the position of the molecule in the simulation box. The coupling of resolutions is achieved through the use of a thermodynamic force in the transition region that compensates for the chemical potential difference between the two resolutions. Another approach is the use of virtual sites to couple the two levels of resolution.[175,176] With the help of these virtual sites, the interactions between CG and atomistic molecules are treated the same way as pure CG–CG interactions, and thus no need for additional parameters arises. For each of these methods, however, applications have so far been limited to simple test systems. The real benefit of multiscale methods has yet to come.

CONCLUSIONS

This review hopefully has provided the reader with a perspective on CG modeling of biosystems. Given the large variety of biomolecular processes, covering many length and time scales, CG models offer access to otherwise unreachable dimensions. However, the universal CG FF does not exist. The real challenge is to choose the right model for the right problem, and to know the inherent limitations. Keeping this in mind, we foresee a bright future for CG modeling, which will claim its place—bridging the microscopic and macroscopic worlds.

REFERENCES

1. Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K. Discovery through the computational microscope. *Structure* 2009, 17:1295–1306.
2. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys* 2012, 41:429–452.
3. Voth GA. *Coarse-Graining of Condensed Phase and Biomolecular Systems*. Boca Raton, FL: CRC Press, Taylor and Francis; 2009.
4. Bond PJ, Holyoake J, Ivetac A, Khalid S, Sansom MSP. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *J Struct Biol* 2007, 157:593–605.
5. Lyubartsev AP, Laaksonen A. Calculation of effective interaction potentials from radial distribution functions: a reverse Monte Carlo approach. *Phys Rev E* 1995, 52:3730.
6. Reith D, Putz M Muller-Plathe F. Deriving effective mesoscale potentials from atomistic simulations. *J Comput Chem* 2003, 24:1624–1636.
7. Izvekov S, Voth GA. A multiscale coarse-graining method for biomolecular systems. *J Phys Chem B* 2005, 109:2469–2473.
8. Noid WG, Chu JW, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J Chem Phys* 2008, 128:244114.
9. Brini E, Algaer EA, Ganguly P, Li CL, Rodriguez-Ropero F, van der Vegt NFA. Systematic coarse-graining methods for soft matter simulations—a review. *Soft Matter* 2013, 9:2108–2119.
10. Rzepiela AJ, Schäfer LV, Goga N, Risselada HJ, de Vries AH, Marrink SJ. Reconstruction of atomistic details from coarse-grained structures. *J Comp Chem* 2010, 31:1333–1343.
11. Stansfeld PJ, Sansom MSP. From coarse grained to atomistic: a serial multiscale approach to membrane protein simulations. *J Chem Theory Comput* 2011, 7:1157–1166.
12. Shehu A, Kavraki LE, Clementi C. Multiscale characterization of protein conformational ensembles. *Proteins* 2009, 76:837–851.

13. Henderson RL. A uniqueness theorem for fluid pair correlation functions. *Phys Lett A* 1974, 49:197.
14. Chaimovich A, Shell MS. Coarse-graining errors and numerical optimization using a relative entropy framework. *J Chem Phys* 2011, 134:094112.
15. Brini E, van der Vegt NFA. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *J Chem Phys* 2012, 137:154113.
16. Noid WG. Systematic methods for structurally consistent coarse-grained models. In: Monticelli L, Salonen E, eds. *Methods in Molecular Biology*. Vol 924. New York: Springer Science+Business Media; 2013, 19.
17. Marrink SJ, de Vries AH, Mark AE. Coarse grained model for semiquantitative lipid simulations. *J Phys Chem B* 2004, 108:750–760.
18. Riniker S, van Gunsteren WF. A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations. *J Chem Phys* 2011, 134:084110.
19. Yesylevskyy SO, Schäfer LV, Sengupta D, Marrink SJ. Polarizable water model for the coarse-grained MARTINI force field. *PLOS Comput Biol* 2010, 6:e1000810.
20. Wu Z, Cui Q, Yethiraj A. A new coarse-grained model for water: the importance of electrostatic interactions. *J Phys Chem B* 2010, 114:10524–10529.
21. Orsi M, Essex JW. The ELBA force field for coarse-grain modeling of lipid membranes. *PLOS Comput Biol* 2011, 6:e28637.
22. Darré L, Machado MR, Dans PD, Herrera FE, Pantano S. Another coarse grain model for aqueous solvation: WAT FOUR? *J Chem Theory Comput* 2010, 6:3793–3807.
23. Brannigan G, Lin LCL, Brown FLH. Implicit solvent simulation models for biomembranes. *Eur Biophys J* 2006, 35:104–124.
24. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. A united-residue force field for off-lattice protein-structure simulations. 1. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem* 1997, 18:849–873.
25. Ouldridge TE, Louis AA, Doye JPK. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys Rev Lett* 2010, 104:178101.
26. Srinivas G, Cheng X, Smith JC. A solvent-free coarse grain model for crystalline and amorphous cellulose fibrils. *J Chem Theory Comput* 2011, 7:2539–2548.
27. Knotts TA, Rathore N, Schwartz DC, de Pablo JJ. A coarse grain model for DNA. *J Chem Phys* 2007, 126:084901.
28. Dans PD, Zeida A, Machado MR, Pantano S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J Chem Theory Comput* 2010, 6:1711–1725.
29. Cao Z, Dama JF, Lu L, Voth GA. Solvent free ionic solution models from multiscale coarse-graining. *J Chem Theory Comput* 2013, 9:172–178.
30. Pizzitutti F, Marchi M, Borgis D. Coarse-graining the accessible surface and the electrostatics of proteins for protein−protein interactions. *J Chem Theory Comput* 2007, 3:1867–1876
31. Hadley KR, McCabe C. Coarse-grained molecular models of water: a review. *Mol Simul* 2012, 38:671–681.
32. Darré L, Machado MR, Pantano S. Coarse-grained models of water. *WIREs Comput Mol Sci* 2012, 2:921–930.
33. Johnson ME, Head-Gordon T, Louis AA. Representability problems for coarse-grained water potentials. *J Chem Phys* 2007, 126:144509–144519.
34. Wang H, Junghans C, Kremer K. Comparative atomistic and coarse-grained study of water: what do we lose by coarse-graining? *Eur Phys J E* 2009, 28:221–229.
35. Hadley KR, McCabe C. On the investigation of coarse-grained models for water: balancing computational efficiency and the retention of structural properties. *J Phys Chem B* 2010, 114:4590–4599.
36. van Hoof B, Markvoort AJ, van Santen RA, Hilbers PAJ. The CUMULUS coarse graining method: transferable potentials for water and solutes. *J Phys Chem B* 2011, 115:10001–10012.
37. He X, Shinoda W, DeVane R, Klein ML. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Mol Phys* 2010, 108:2007–2020.
38. Chiu SW, Scott HL, Jakobsso E. A coarse-grained model based on Morse potential for water and n-alkanes. *J Chem Theory Comput* 2010, 6:851–863.
39. Molinero V, Goddard III VA. M3B: a coarse grain force field for molecular simulations of malto-oligosaccharides and their water mixtures. *J Phys Chem B* 2004, 108:1414–1427.
40. Shelley JC, Shelley MY, Reeder RC, Bandyopadhyay S, Klein ML. A coarse grained model for phospholipid simulations. *J Phys Chem B* 2001, 105:4464–4470.
41. Shinoda, W, DeVane, R, Klein, ML. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Mol Simulat* 2007, 33:27–36.
42. Molinero V, Moore EB. Water modeled as an intermediate element between carbon and silicon. *J Phys Chem B* 2009, 113:4008–4016.
43. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* 2007, 111:7812–7824.
44. DeMille RC, Molinero V. Coarse-grained ions without charges: reproducing the solvation structure of

NaCl in water using short-ranged potentials. *J Chem Phys* 2009, 131:034107.

45. Marrink SJ, Mark AE. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J Am Chem Soc* 2003, 125:15233–15242.
46. Ayton GS, Voth GA. Hybrid coarse-graining approach for lipid bilayers at large length and time scales. *J Phys Chem B* 2009, 113:4413–4424.
47. Ha-Duong T, Basdevant N, Borgis D. A polarizable coarse-grained water model for coarse-grained proteins simulations. *Chem Phys Lett* 2009, 468:79–82.
48. Darré L, Tek A, Baaden M, Pantano S. Mixing atomistic and coarse grain solvation models for MD simulations: let WT4 handle the bulk. *J Chem Theory Comput* 2012, 8:3880–3894.
49. Masella M, Borgis D, Cuniasse P. Combining a polarizable force-field and a coarse-grained polarizable solvent model. II. Accounting for hydrophobic effects. *J Comput Chem* 2011, 32:2664–2678.
50. Smit B, Hilbers PAJ, Esselink K, Rupert LAM, Van Os NM, Schlijper AG. Computer simulations of a water/oil interface in the presence of micelles. *Nature* 1990, 348:624–625.
51. Müller M, Katsov K, Schick M. Biological and synthetic membranes: what can be learned from a coarse-grained description? *Phys Rep* 2006, 434:113–176.
52. Venturoli M, Sperotto MM, Kranenburg M, Smit B. Mesoscopic models of biological membranes. *Phys Rep* 2006, 437:1–54.
53. Bennun SV, Hoopes MI, Xing C, Faller R. Coarse-grained modeling of lipids. *Chem Phys Lipids* 2009, 159:59–66.
54. Lyubartsev AP, Rabinovich AL. Recent development in computer simulations of lipid bilayers. *Soft Matter* 2011, 7:25–39.
55. Shinoda W, DeVane R, Klein ML. Zwitterionic lipid assemblies: molecular dynamics studies of monolayers, bilayers, and vesicles using a new coarse grain force field. *J Phys Chem B* 2010, 114:6836–6849.
56. Jusufi A, DeVane, RH, Shinoda W, Klein ML. Nanoscale carbon particles and the stability of lipid bilayers. *Soft Matter* 2011, 7:1139–1146.
57. Marrink SJ, de Vries AH, Harroun TA, Katsaras J, Wassall SR. Cholesterol shows preference for the interior of polyunsaturated lipid membranes. *J Am Chem Soc* 2008, 130:10–11.
58. Bulacu M, Périole X, Marrink SJ. In silico design of robust bolalipid membranes. *Biomacromolecules* 2012, 13:196–205.
59. Risselada HJ, Marrink SJ. The molecular face of lipid rafts in model membranes. *Proc Natl Acad Sci USA* 2008, 105:17367–17372.
60. Baoukina S, Marrink SJ, Tieleman DP. Molecular structure of membrane tethers *Biophys J* 2012, 102:1866–1871.
61. Orsi M, Haubertin DY, Sanderson WE, Essex JW. A quantitative coarse-grain model for lipid bilayers. *J Phys Chem B* 2008, 112:802–815.
62. Michel J, Orsi M, Essex JW. Prediction of partition coefficients by multiscale hybrid atomic level/coarse-grain simulations. *J Phys Chem B* 2008, 112:657–660.
63. Izvekov S, Voth GA. A multiscale coarse-graining method for biomolecular systems. *J Phys Chem B* 2005, 109:2469–2473.
64. Izvekov S, Voth GA. Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *J Chem Theory Comput* 2006, 2:637–648.
65. Lu L, Voth GA. Systematic coarse-graining of a multicomponent lipid bilayer. *J Phys Chem B* 2009, 113:1501–1510.
66. Kranenburg M, Nicolas J-P, Smit B. Comparison of mesoscopic phospholipid-water models. *Phys Chem Chem Phys* 2004, 6:4142–4151.
67. de Meyer F, Smit B. Effect of cholesterol on the structure of a phospholipid bilayer. *Proc Natl Acad Sci USA* 2009, 106:3654–3658.
68. Kranenburg M, Smit B. Phase behavior of model lipid bilayers. *J Phys Chem B* 2005, 109:6553–6563.
69. Rodgers JM, Sørensen J, de Meyer FJM, Schiøtt B, Smit B. Understanding the phase behavior of coarse-grained model lipid bilayers through computational calorimetry. *J Phys Chem B* 2012, 116:1551–1569.
70. Lyubartsev AP. Multiscale modeling of lipids and lipid bilayers. *Eur Biophys J* 2005, 35:53–61.
71. Wang ZJ, Deserno M. A systematically coarse-grained solvent-free model for quantitative phospholipid bilayer simulations. *J Phys Chem B* 2010, 114:11207–11220.
72. Sodt AJ, Head-Gordon T. An implicit solvent coarse-grained lipid model with correct stress profile. *J Chem Phys* 2010, 132:205103.
73. Curtis, EM, Hall, CK. Molecular dynamics simulations of DPPC bilayers using "LIME", a new coarse-grained model. *J Phys Chem B* 2013, 117:5019–5030.
74. Srivastava A, Voth GA. Hybrid approach for highly coarse-grained lipid bilayer models. *J Chem Theory Comput* 2013, 9:750–765.
75. Levitt M, Warshel A. Computer-simulation of protein folding. *Nature* 1975, 253:694–698.
76. Dill KA. Theory for the folding and stability of globular proteins. *Biochemistry* 1985, 24:1501–1509.
77. Koga N, Takada S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J Mol Biol* 2001, 313:171–180.
78. Levy Y, Wolynes PG, Onuchic JN. Protein topology determines binding mechanism. *Proc Natl Acad Sci USA* 2004, 101:511–516.

79. Head-Gordon T, Brown S. Minimalist models for protein folding and design. *Curr Opin Struct Biol* 2003, 13:160–167.
80. Hills RD Jr, Brooks CL III. Insights from coarse-grained gō models for protein folding and dynamics. *Int J Mol Sci* 2009, 10:889–905.
81. Bereau T, Deserno M. Generic coarse-grained model for protein folding and aggregation. *J Chem Phys* 2009, 130:235106.
82. Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. *Proteins* 2007, 69:394–408.
83. Chebaro Y, Pasquali S, Derreumaux P. The coarse-grained OPEP force field for non-amyloid and amyloid proteins. *J Phys Chem B* 2012, 116: 8741–8752.
84. Monticelli L, Kandasamy KS, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse grained force field: extension to proteins. *J Chem Theory Comput* 2008, 4:819–834.
85. Miyazawa S, Jernigan RL. Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 1996, 256:623–644.
86. Bereau T, Bachmann M, Deserno M. Interplay between secondary and tertiary structure formation in protein folding cooperativity. *J Am Chem Soc* 2010, 132:13129–13131.
87. Barducci A, Bonomi M, Derreumaux P. Assessing the quality of the OPEP coarse-grained force field. *J Chem Theory Comput* 2011, 7:1928–1934.
88. Maupetit J, Derreumaux P, Tuffery P. A fast method for large-scale de novo peptide and miniprotein structure prediction. *J Comput Chem* 2010, 31:726–738.
89. Wei G, Mousseau N, Derreumaux P. Computational simulations of the early steps of protein aggregation. *Prion* 2007, 1:3–8.
90. Chebaro Y, Dong X, Laghaei R, Derreumaux P, Mousseau N. Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *J Phys Chem B* 2009, 113:267–274.
91. Periole X, Cavalli M, Marrink SJ, Ceruso MA. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *J Chem Theory Comput* 2009, 5:2531–2543.
92. Singh G, Tieleman DP. Using the Wimley–White hydrophobicity scale as a direct quantitative test of force fields: the Martini coarse-grained model. *J Chem Theory Comput* 2011, 7:2316–2324.
93. de Jong DH, Periole X, Marrink SJ. Dimerization of amino acid side chains: lessons from the comparison of different force fields. *J Chem Theory Comput* 2012, 8:1003–1014.
94. de Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schäfer LV, Periole X, Tieleman DP, Marrink SJ. Improved parameters for the Martini coarse-grained protein force field. *J Chem Theory Comput* 2013, 9:687–697.
95. Louhivuori M, Risselada HJ, van der Giessen E, Marrink SJ. Release of content through mechano-sensitive gates in pressurized liposomes. *Proc Natl Acad Sci USA* 2010, 107:19856–19860.
96. Periole X, Huber T, Marrink SJ, Sakmar TP. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J Am Chem Soc* 2007, 129:10126–10132.
97. Periole X, Knepp AM, Sakmar TP, Marrink SJ, Huber T. Structural determinants of the supra-molecular organization of G protein-coupled receptors in bilayers. *J Am Chem Soc* 2012, 134:10959–10965.
98. Christoph G, Venkatramanan K, Markus D, Christine P. Optimization of an elastic network augmented coarse grained model to study CCMV capsid deformation. *PLoS ONE* 2013, 8:e60582.
99. Schäfer LV, de Jong DH, Holt A, Rzepiela AJ, de Vries AH, Poolman B, Killian JA, Marrink SJ. Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes. *Proc Natl Acad Sci USA* 2011, 108:1343–1348.
100. Arnarez C, Marrink SJ, Periole X. Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. *Sci Rep* 2013, 3:1263.
101. Liwo A, He Y, Scheraga HA. Coarse-grained force field: general folding theory. *Phys Chem Chem Phys* 2011, 13:16890–16901.
102. Liwo A, Wawak RJ, Scheraga HA, Pincus MR, Rackovsky S. Prediction of protein conformation on the basis of a search for compact structures: test on avian pancreatic polypeptide. *Protein Sci* 1993, 2:1715–1731.
103. Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *J Phys Chem A* 2010, 114: 4471–4485.
104. Oldziej S, Czaplewski C, Liwo A, Chinchio M, Nanias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, et al. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. *Proc Natl Acad Sci USA* 2005, 102:7547–7552.
105. He Y, Liwo A, Weinstein H, Scheraga HA. PDZ binding to the BAR domain of PICK1 is elucidated by coarse-grained molecular dynamics. *J Mol Biol* 2011, 405:298–314.
106. Rojas A, Liwo A, Browne D, Scheraga HA. Mechanism of fiber assembly: treatment of a beta peptide aggregation with a coarse-grained united-residue force field. *J Mol Biol* 2010, 404:537–552.

107. Basdevant N, Borgis D, Ha-Duong T. A coarse-grained protein-protein potential derived from an all-atom force field. *J Phys Chem B* 2007, 111:9390–9399.
108. Ha-Duong T, Basdevant N, Borgis D. Modeling protein–protein recognition in solution using the coarse-grained force field SCORPION. *J Chem Theory Comput* 2013, 9:803–813.
109. Pasi M, Lavery R, Ceres N. PaLaCe: a coarse-grain protein model for studying mechanical properties. *J Chem Theory Comput* 2012, 9:785–793.
110. Kar P, Gopal SM, Cheng YM, Predeus A, Feig M. PRIMO: a transferable coarse-grained force field for proteins. *J Chem Theory Comput* 2013, 9:3769–3788.
111. Gopal SM, Mukherjee S, Yi Ming C, Feig M. PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins* 2010, 78:1266–1281.
112. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J Phys Chem B* 2012, 116:8494–8503.
113. Weihua Z, Schafer NP, Davtyan A, Papoian GA, Wolynes PG. Predictive energy landscapes for protein–protein association. *Proc Natl Acad Sci USA* 2012, 109:19244–19249.
114. Laing C, Schlick T. Computational approaches to RNA structure prediction, analysis, and design. *Curr Opin Struct Biol* 2011, 21:306–318.
115. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* 2008, 14:1164–1173.
116. Xia Z, Gardner DP, Gutell RR, Ren P. Coarse-grained model for simulation of RNA three-dimensional structures. *J Phys Chem B* 2010, 114:13497–13506.
117. Tan RKZ, Harvey SC. Molecular mechanics model of supercoiled DNA. *J Mol Biol* 1989, 205:573–591.
118. Schlick T, Olson WK. Supercoiled DNA energetics and dynamics by computer simulation. *J Mol Biol* 1992, 223:1089–1119.
119. Drukker K, Wu G, Schatz GC. Model simulations of DNA denaturation dynamics. *J Chem Phys* 2001, 114:579–590.
120. de Pablo JJ. Coarse-grained simulations of macromolecules: from DNA to nanocomposites. *Annu Rev Phys Chem* 2011, 62:555–574.
121. Potoyan DA, Savelyev A, Papoian GA. Recent successes in coarse-grained modeling of DNA. *WIREs Comput Mol Sci* 2013, 3:69–83.
122. Hyeon C, Thirumalai D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun* 2011, 2:487.
123. Sambriski EJ, Schwartz DC, de Pablo JJ. A mesoscale model of DNA and its renaturation. *Biophys J* 2009, 96:1675–1690.
124. Freeman GS, Hinckley DM, de Pablo JJ. A coarse-grain three-site-per-nucleotide model for DNA with explicit ions. *J Chem Phys* 2011, 135:165104.
125. DeMille RC, Cheatham TE III, Molinero V. A coarse-grained model of DNA with explicit solvation by water and ions. *J Phys Chem B* 2011, 115:132–142.
126. De Biase PM, Solano CJF, Markosyan S, Czapla L, Noskov SY. BROMOC-D: Brownian dynamics/Monte-Carlo program suite to study ion and DNA permeation in nanopores. *J Chem Theory Comput* 2012, 8:2540–2551.
127. Hinckley DM, Freeman GS, Whitmer JK, de Pablo JJ. 3SPN.2: an experimentally–informed coarse–grained model of DNA. *J Chem Phys* (submitted).
128. Ouldridge TE, Louis AA, Doye JPK. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J Chem Phys* 2011, 134:085101.
129. Šulc P, Romano F, Ouldridge TE, Rovigatti L, Doye JPK, Louis AA. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J Chem Phys* 2012, 137:135101.
130. Pasquali S, Derreumaux P. HiRE-RNA: a high resolution coarse-grained energy model for RNA. *J Phys Chem B* 2010, 114:11957–11966.
131. Linak MC, Tourdot R, Dorfman KD. Moving beyond Watson–Crick models of coarse grained DNA dynamics. *J Chem Phys* 2011, 135:205102.
132. Morriss-Andrews A, Rottler J, Plotkin SS. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J Chem Phys* 2010, 132:035105.
133. Bush CA, Martin-Pastor M, Imberty A. Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides. *Annu Rev Biophys Biomol Struct* 1999, 28:269–293.
134. Duus J, Gotfredsen CH, Bock K. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem Rev* 2000, 100:4589–4614.
135. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA. Conformational studies of oligosaccharides and glycopeptides: complementarity of NMR, X-ray crystallography, and molecular modelling. *Chem Rev* 2002, 102:371–386.
136. Yu F, Prestegard J. Structural monitoring of oligosaccharides through 13C enrichment and NMR observation of acetyl groups. *Biophys J* 2006, 91:1952–1959.

137. Liu P, Izvekov S, Voth GA. Multiscale coarse-graining of monosaccharides. *J Phys Chem B* 2007, 111:11566–11575.
138. López CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hünenberger PH, Marrink SJ. Martini coarse-grained force field: extension to carbohydrates. *J Chem Theory Comput* 2009, 5:3195–3210.
139. Wohlert J, Berglund LA. A coarse-grained model for molecular dynamics simulations of native cellulose. *J Chem Theory Comput* 2011, 7:753–760.
140. López CA, Sovova Z, van Eerden FJ, de Vries AH, Marrink SJ. Martini force field parameters for glycolipids. *J Chem Theory Comput* 2013, 9:1694–1708.
141. López CA, de Vries AH, Marrink SJ. Computational microscopy of cyclodextrin mediated cholesterol extraction from lipid model membranes. *Sci Rep* 2013, 3:2071.
142. Bellesia G, Chundawat SPS, Langan P, Redondo A, Dale BE, Gnanakaran S. Coarse-grained model for the interconversion between native and liquid ammonia-treated crystalline cellulose. *J Phys Chem B* 2012, 116:8031–8037.
143. Bathe M, Rutledge GC, Grodzinsky AJ, Tidor B. A coarse-grained molecular model for glycosaminoglycans: application to chondroitin, chondroitin sulfate, and hyaluronic acid. *Biophys J* 2005, 88:3870–3887.
144. Sattelle BM, Shakeri J, Almond A. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* 2013, 14:1149–1159.
145. Markutsya S, Devarajan A, Baluyut JY, Windus TL, Gordon MS, Lamm MA. Evaluation of coarse-grained mapping schemes for polysaccharide chains in cellulose. *J Chem Phys* 2013, 138:214108.
146. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 1998, 282:740–744.
147. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science* 2011, 334:517–520.
148. Snow CD, Nguyen H, Pande VS, Gruebele M. Absolute comparison of simulated and experiment protein-folding dynamics. *Nature* 2002, 420:102–106.
149. Liwo A, Khalili M, Scheraga HA. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc Natl Acad Sci USA* 2005, 102:2362–2367.
150. Rojas AV, Liwo A, Scheraga HA. Molecular dynamics with the united-residue force field: ab initio folding simulations of multichain proteins. *J Phys Chem B* 2007, 111:293–309.
151. Sukharev SI, Blount P, Martinac B, Blattner FR, Kung C. A large-conductance mechanosensitive channel in E. *coli* encoded by mscL alone. *Nature* 1994, 368:265–268.
152. Steinbacher S, Bass R, Strop P, Rees DC. Structures of the prokaryotic mechanosensitive channels MscL and MscS. *Curr Top Membr* 2007, 58:1–24.
153. Gullingsrud J, Kosztin D, Schulten K. Structural deter- minants of MscL gating studied by molecular dynamics simulations. *Biophys J* 2001, 80:2074–2081.
154. Elmore DE, Dougherty DA. Molecular dynamics simulations of wild-type and mutant forms of the M*ycobacterium tuberculosis* MscL channel. *Biophys J* 2001, 81:1345–1359.
155. Rui H, Kumar R, Im W. Membrane tension, lipid adaptation, conformational changes, and energetics in MscL gating. *Biophys J* 2011, 101:671–679.
156. Yefimov S, van der Giessen E, Onck PR, Marrink SJ. Mechanosensitive membrane channels in action. *Biophys J* 2008, 94:2994–3002.
157. Ollila OHS, Louhivuori M, Marrink SJ, Vattulainen I. Protein shape change has a major effect on the gating energy of a mechanosensitive channel. *Biophys J* 2011, 100:1651–1659.
158. Koçer A, Walko M, Meijberg W, Feringa BL. A light-actuated nanovalve derived from a channel protein. *Science* 2005, 309:755–758.
159. Domanski J, Marrink SJ, Schäfer LV. Transmembrane helices can induce domain formation in crowded model biomembranes. *Biochim Biophys Acta—Biomembr* 2012, 1818:984–994.
160. Sambriski EJ, Schwartz DC, de Pablo JJ. Uncovering pathways in DNA oligonucleotide hybridization via transition state analysis. *Proc Natl Acad Sci USA* 2009, 106:18125–18130.
161. Deserno M. Mesoscopic membrane physics: concepts, simulations, and selected applications. *Macromol Rapid Commun* 2009, 30:752–771.
162. Mullinax JW, Noid WG. Extended ensemble approach for deriving transferable coarse-grained potentials. *J Chem Phys* 2009, 131:104110.
163. Rühle V, Junghans C, Lukyanov A, Kremer K, Andrienko D. Versatile object-oriented toolkit for coarse-graining applications. *J Chem Theory Comput* 2009, 5:3211–3223.
164. Mirzoev A, Lyubartsev AP. MagiC: software package for multiscale modeling. *J Chem Theory Comput* 2013, 9:1512–1520.
165. Zhang ZY, Lu LY, Noid WG, Krishna V, Pfaendtner J, Voth GA. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophys J* 2008, 95:5073–5083.
166. Nielsen S, Bulo R, Moore P, Ensing B. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Phys Chem Chem Phys* 2010, 12:12401–12414.
167. Kamerlin SCL, Warshel A. Multiscale modeling of biological functions. *Phys Chem Chem Phys* 2011, 13:10401–10411.

168. Ayton GS, Noid WG, Voth GA. Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr Opin Struct Biol* 2007, 17:192–198.
169. Peter C, Kremer K. Multiscale simulation of soft matter systems—from the atomistic to the coarse-grained level and back. *Soft Matter* 2009, 5:4357–4366.
170. Meier K, Choutko A, Dolenc J, Eichenberger AP, Riniker S, van Gunsteren WF. Multi-resolution simulation of biomolecular systems: a review of methodological issues. *Angew Chem Int Ed* 2013, 52:2820–2834.
171. Murtola T, Bunker A, Vattulainen I, Deserno M, Karttunen M. Multiscale modeling of emergent materials: biological and soft matter. *Phys Chem Chem Phys* 2009, 11:1869–1892.
172. Han W, Schulten K. Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: improved backbone hydration and interactions between charged side chains. *J Chem Theory Comput* 2012, 8:4413–4424.
173. Neri M, Anselmi C, Cascalla M, Maritan A, Carloni P. Coarse-grained model of proteins incorporating atomistic detail of the active site. *Phys Rev Lett* 2005, 95: 218102.
174. Praprotnik M, Delle Site L, Kremer K. Adaptive resolution molecular-dynamics simulation: changing the degrees of freedom on the fly. *J Chem Phys* 2005, 123:224106.
175. Rzepiela AJ, Louhivuori M, Peter C, Marrink SJ. Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. *Phys Chem Chem Phys* 2011, 13:10437–10448.
176. Wassenaar TA, Ingólfsson HI, Prieß M, Marrink SJ, Schäfer LV. Mixing martini: electrostatic coupling in hybrid atomistic—coarse-grained biomolecular simulations. *J Phys Chem B* 2013, 117: 3516–3530.

FURTHER READING/RESOURCES

For additional information and updates on the models, please see the model or developing groups website. Some sites also provide help with simulation setup, tutorials, and downloadable FFs.

Bereau and Deserno	cmu.edu/biolphys/deserno
de Pablo et al.	molecularengineering.uchicago.edu/people/juan-de-pablo
ELBA	www.orsi.sems.qmul.ac.uk/elba
Klein CG models	icms.cst.temple.edu
Martini	cgmartini.nl
OPEP	www-lbt.ibpc.fr/LBT
Ouldridge et al.	dna.physics.ox.ac.uk
PaLeCe	www.ibcp.fr/_Richard-LAVERY
PRIMO	feig.bch.msu.edu
SIRAH	www.sirahff.com
Smit's DPD model	www.cchem.berkeley.edu/molsim
UNRES	unres.pl
Voth CG models	vothgroup.uchicago.edu

